

# Extreme Dependence Models

Boris Béranger <sup>\*†</sup>, Simone A. Padoan <sup>‡</sup>

## Abstract

Extreme values of real phenomena are events that occur with low frequency, but can have a large impact on real life. These are, in many practical problems, high-dimensional by nature (e.g. Tawn, 1990; Coles and Tawn, 1991). To study these events is of fundamental importance. For this purpose, probabilistic models and statistical methods are in high demand. There are several approaches to modelling multivariate extremes as described in Falk et al. (2011), linked to some extent. We describe an approach for deriving multivariate extreme value models and we illustrate the main features of some flexible extremal dependence models. We compare them by showing their utility with a real data application, in particular analyzing the extremal dependence among several pollutants recorded in the city of Leeds, UK.

## 1 Introduction

Statistical analyses of extreme events are of crucial importance for risk assessment in many areas such as the financial market, telecommunications, industry, environment and health. For example governments and insurance companies need to statistically quantify the frequency of natural disasters in order to plan risk management and take preventive actions.

Several examples of univariate analysis are available, for instance in Coles (2001). Two main approaches are used in applications, the block-maximum and the peak over a threshold. These are based on the generalized extreme value (GEV) distribution and the generalized Pareto distribution

---

<sup>\*</sup>Theoretical and Applied Statistics Laboratory (LSTA), University Pierre and Marie Curie - Paris 6, F-75005, Paris, France

<sup>†</sup>School of Mathematics and Statistics, University of New South Wales, Sydney, Australia

<sup>‡</sup>Department of Decision Sciences, Bocconi University of Milan, 20136 Milano, Italy

(GPD), which are milestones of the extreme value theory, see e.g. Coles (2001, Ch. 3–4) and the references therein.

Many practical problems in finance, the environment, etc. are high-dimensional by nature, for example when analyzing the air quality in an area, the amount of pollution depends on the levels of different pollutants and the interaction between them. Today the extreme value theory provides a sufficiently mature framework for applications in the multivariate case. Indeed a large number of theoretical results and statistical methods and models are available, see for instance the monographs Resnick (2007), de Haan and Ferreira (2006), Falk et al. (2011), Beirlant et al. (2006), Coles (2001) and Kotz and Nadarajah (2000). In this article we review some basic theoretical results on the extreme values of multivariate variables (multivariate extremes for brevity). With the block-maximum approach we explain what type of dependence structures can be described. We discuss the main features of some families of parametric extremal dependence models. By means of real data analysis we show the utility of these extremal dependence models when assessing the dependence of multivariate extremes. Their utility is also illustrated when estimating the probabilities that multivariate extreme events occur.

The analysis of real phenomena such as heavy rainfall, heat waves and so on is a challenging task. The first difficulty is the complexity of the data, i.e. observations are collected over space and time. In this case, theory deals with extremes of temporal- or spatial-processes (e.g. de Haan and Ferreira, 2006, Ch. 9). Examples of such statistical analysis are Davison et al. (2012), Davison and Gholamrezaee (2012), for a simple review see Padoan (2013a). This theory is closely linked to that of multivariate extremes presented here. The second difficulty is that the dependence of multivariate extremes is not always well captured by the models illustrated here. Ledford and Tawn (1996, 1997) have shown that in some applications a more suitable dependence structure is described by the so called *asymptotic independence*. This framework has been recently extended to continuous processes (e.g. De Haan and Zhou, 2011; Wadsworth and Tawn, 2012; Padoan, 2013c). These motivations make the multivariate extreme value theory a very active research field at present.

The paper is organized as follows. In Section 1.2 a definition of multivariate extremes is provided and the main characteristics are presented. In Section 1.3 some of the most popular extremal dependence models are described. In Section 1.4 some estimation methods are discussed and in Section 1.5 the analysis of the extremes of multiple pollutants is performed.

## 2 Multivariate Extremes

Applying the block-maximum approach to every component of a multivariate random vector gives rise to a definition of multivariate extremes. Specifically, for  $d \in \mathbb{N}$ , let  $I = \{1, \dots, d\}$  be an index set and  $\mathbf{X} = (X_1, \dots, X_d)$  be an  $\mathbb{R}^d$ -valued random vector with joint (probability) distribution function  $F$  and marginal distribution functions  $F_j = F(\infty, \dots, x_j, \dots, \infty)$ ,  $j \in I$ . Suppose that  $\mathbf{X}_1, \dots, \mathbf{X}_n$  are  $n$  independent and identically distributed (i.i.d.) copies of  $\mathbf{X}$ . The sample vector of componentwise maxima (sample maxima for brevity) is  $\mathbf{M}_n = (M_{n,1}, \dots, M_{n,d})$ , where  $M_{n,j} = \max(X_{1,j}, \dots, X_{n,j})$ .

Typically, in applications the distribution  $F$  is unknown and so the distribution of the sample maxima is also unknown. A possible solution is to study the asymptotic distribution of  $M_n$  as  $n \rightarrow \infty$  and to use it as an approximation for a large but finite sample size, resulting in an approximate distribution for multivariate extremes. At a first glance, this notion of multivariate extremes may seem too simple to provide a useful approach for applications. However, a number of theoretical results justify its practical use. For example, with this definition of multivariate extremes, the dependence that arises is linked to the dependence that all the components of  $\mathbf{X}$  are simultaneously large. Thus, by estimating these dependence structures we are also able to estimate the probabilities that multiple exceedances occur.

### 2.1 Multivariate extreme value distributions

The asymptotic distribution of  $\mathbf{M}_n$  is derived with a similar approach to the univariate case. Assume there are sequences of normalizing constants  $\mathbf{a}_n = (a_{n1}, \dots, a_{nd}) > \mathbf{0}$ , with  $\mathbf{0} = (0, \dots, 0)$ , and  $\mathbf{b}_n = (b_{n1}, \dots, b_{nd}) \in \mathbb{R}^d$  such that

$$\text{pr}\left(\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{x}\right) = F^n(\mathbf{a}_n \mathbf{x} + \mathbf{b}_n) \rightarrow G(\mathbf{x}), \quad n \rightarrow \infty, \quad (1)$$

for all the continuity points  $\mathbf{x}$  of a non-degenerate distribution  $G$ . The class of the limiting distributions in (1) is called *multivariate extreme value distributions* (MEVDs) (Resnick, 2007, p. 263). A distribution function  $F$  that satisfies the convergence result (1) is said to be in the (*maximum*) *domain of attraction* of  $G$  (de Haan and Ferreira, 2006, pp. 226–229). An attractive property of MEVDs is the *max-stability*. A distribution  $G$  on  $\mathbb{R}^d$  is max-stable if for every  $n \in \mathbb{N}$ , there exists

sequences  $\mathbf{a}_n > \mathbf{0}$  and  $\mathbf{b}_n \in \mathbb{R}^d$  such that

$$G(\mathbf{a}_n \mathbf{x} + \mathbf{b}_n) = G^{1/n}(\mathbf{x}), \quad (2)$$

(Resnick, 2007, Proposition 5.9). As a consequence,  $G$  is such that  $G^a$  is a distribution for every  $a > 0$ . A class of distributions that satisfies such a property is named *max-infinitely divisible* (max-id). More precisely, a distribution  $G$  on  $\mathbb{R}^d$  is max-id, if for any  $n \in \mathbb{N}$  there exists a distribution  $F_n$  such that  $G = F_n^n$  (Resnick, 2007, p. 252). This means that  $G$  can always be defined through the distribution of the sample maxima of  $n$  i.i.d. random vectors.

In order to characterize the class of MEVDs we need to specify: a) the form of the marginal distributions, b) the form of the dependence structure.

a) To illustrate the first feature is fairly straightforward. If  $F$  converges, then so too does the marginal distributions  $F_j$  for all  $j \in I$ . Choosing  $a_{jn}$  and  $b_{jn}$  for all  $j \in I$  as in de Haan and Ferreira (2006, Corollary 1.2.4), implies that each marginal distribution of  $G$  is a generalized extreme value (GEV), i.e.

$$G(\infty, \dots, x_j, \dots, \infty) = \exp \left[ - \left\{ 1 + \xi_j \left( \frac{x_j - \mu_j}{\sigma_j} \right) \right\}_+^{-1/\xi_j} \right], \quad j \in I,$$

where  $(x)_+ = \max(0, x)$ ,  $-\infty < \mu_j, \xi_j < \infty$ ,  $\sigma_j > 0$  (de Haan and Ferreira, 2006, pp. 208–211). Because the marginal distributions are continuous then  $G$  is also continuous.

b) The explanation of the dependence form is more elaborate, although it is not complicated. The explanation is based on three steps: 1)  $G$  is transformed so that its marginal distributions are equal, 2) a Poisson point process (PPP) is used to represent the standardised distribution, 3) the dependence form is made explicit by means of a change of coordinates. Here are the steps.

1) Let  $U_j(a) = F_j^{\leftarrow}(1 - 1/a)$ , with  $a > 1$ , be the left-continuous inverse of  $F_j$ , for all  $j \in I$ . The sequences  $a_{nj}$  and  $b_{nj}$  in (1) are such that for all  $y_j > 0$ ,

$$\lim_{n \rightarrow \infty} \frac{U_j(ny_j) - b_n}{a_n} = \frac{\sigma_j(y_j^{\xi_j} - 1)}{\xi_j} + \mu_j, \quad j \in I,$$

and therefore

$$\begin{aligned} & \lim_{n \rightarrow \infty} F^n\{U_1(ny_1), \dots, U_d(ny_d)\} \\ &= G \left( \frac{\sigma_1(y_1^{\xi_1} - 1)}{\xi_1} + \mu_1, \dots, \frac{\sigma_d(y_d^{\xi_d} - 1)}{\xi_d} + \mu_d \right) \equiv G_0(\mathbf{y}), \end{aligned} \quad (3)$$

for all continuity points  $\mathbf{y} > \mathbf{0}$  of  $G_0$  (see de Haan and Ferreira, 2006, Theorems 1.1.6, 6.1.1).  $G_0$  is a MEVD with identical unit Fréchet marginal distributions.

Now, for all  $\mathbf{y} > \mathbf{0}$  such that  $0 < G_0(\mathbf{y}) < 1$ , by taking the logarithm on the right and left side of (3) and using a first order Taylor expansion of  $\log F\{U_1(ny_1), \dots, U_d(ny_d)\}$ , as  $n \rightarrow \infty$ , it follows that

$$\lim_{n \rightarrow \infty} n[1 - F\{U_1(ny_1), \dots, U_d(ny_d)\}] = -\log G_0(\mathbf{y}) \equiv V(\mathbf{y}). \quad (4)$$

The function  $V$ , named *exponent (dependence) function*, represents the dependence structure of multiple extremes (extremal dependence for brevity). According to (4) the derivation of  $V$  depends on the functional form of  $F$ . In most of the practical problems the latter is unknown. A possible solution is obtained exploiting the max-id property of  $G_0$ , which says that every max-id distribution permits a PPP representation, see Resnick (2007, pp. 257–262) and Falk et al. (2011, pp. 141–142).

2) Let  $N_n(\cdot)$  be a PPP defined by

$$N_n(\mathcal{A}) := \sum_{i=1}^{\infty} \mathbb{I}_{\{\mathbf{P}_i\}}(\mathcal{A}), \quad \mathbb{I}_{\{\mathbf{P}_i\}}(\mathcal{A}) = \begin{cases} 1, & \mathbf{P}_i \in \mathcal{A}, \\ 0, & \mathbf{P}_i \notin \mathcal{A}, \end{cases}$$

where  $\mathcal{A} \subset \mathbb{A}$  with  $\mathbb{A} := (0, \infty) \times \mathbb{R}_+^d$ ,

$$\mathbf{P}_i = \left[ \frac{i}{n}, \left\{ 1 + \xi_1 \left( \frac{X_{i1} - b_{n1}}{a_{n1}} \right) \right\}^{\frac{1}{\xi_1}}, \dots, \left\{ 1 + \xi_d \left( \frac{X_{id} - b_{nd}}{a_{nd}} \right) \right\}^{\frac{1}{\xi_d}} \right],$$

for every  $n \in \mathbb{N}$  and  $\mathbf{X}_i$ ,  $i = 1, 2, \dots$  are i.i.d random vectors with distribution  $F$ . The intensity measure is  $\zeta \times \eta_n$  where  $\zeta$  is the Lebesgue measure and for every  $n \in \mathbb{N}$  and all critical regions defined by  $\mathcal{B}_{\mathbf{y}} := \mathbb{R}_+^d \setminus [\mathbf{0}, \mathbf{y}]$  with  $\mathbf{y} > \mathbf{0}$ ,

$$\eta_n(\mathcal{B}_{\mathbf{y}}) = n[1 - F\{U_1(ny_1), \dots, U_d(ny_d)\}],$$

is a finite measure. If the limit in (3) holds, then  $N_n$  converges weakly to  $N$  as  $n \rightarrow \infty$ , i.e. a PPP with intensity measure  $\zeta \times \eta$  where

$$\eta(\mathcal{B}_{\mathbf{y}}) = \eta\{(\mathbf{v} \in \mathbb{R}_+^d : v_1 > y_1 \text{ or } \dots \text{ or } v_d > y_d)\} \equiv V(\mathbf{y}), \quad \mathbf{y} > \mathbf{0},$$

is a fine measure, named *exponent measure* (see de Haan and Ferreira, 2006, Theorems 6.1.5, 6.1.11). Observe that  $\eta$  must concentrate on  $\overline{\mathbb{R}} = \mathbb{R}_+^d \setminus \{\mathbf{0}\}$  in order to be uniquely determined. Also,  $\eta$  must

satisfy  $\eta(\infty) = 0$ , see Falk et al. (2011, p. 143) for details.

This essentially means that numbering the rescaled observations that fall in a critical region, e.g. see the shaded sets in the left panels of Figure 1, where at least one coordinate is large, makes it possible for (3) to be computed using the void probability of  $N$ , that is

$$\begin{aligned} G_0(\mathbf{y}) &= \text{pr}[N\{(0, 1] \times \mathcal{B}_{\mathbf{y}}\} = 0] \\ &= \exp(-[\zeta\{(0, 1]\} \times \eta(\mathcal{B}_{\mathbf{y}})]) \\ &= \exp\{-V(\mathbf{y})\} \quad \mathbf{y} > \mathbf{0}. \end{aligned} \tag{5}$$

From Figure 1 we see that in the case of strong dependence (top-left panel) all the coordinates of the extremes are large, while in the case of weak dependence (bottom-left panels) only one coordinate of the extremes is large.

At this time it remains to be specify the structure of the exponent measure. This task is simpler to fulfil when working with pseudo-polar coordinates.

3) With unit Fréchet margins, the stability property (2) can be rephrased by  $G_0^a(a\mathbf{y}) = G_0(\mathbf{y})$  for any  $a > 0$ , implying that  $\eta$  satisfies the homogeneity property

$$\eta(a\mathcal{B}_{\mathbf{y}}) = \eta(\mathcal{B}_{\mathbf{y}})/a, \tag{6}$$

for all  $\mathcal{B}_{\mathbf{y}} \subset \overline{\mathbb{R}}$ , where  $\mathcal{B}_{\mathbf{y}} := \overline{\mathbb{R}} \setminus (\mathbf{0}, \mathbf{y}]$  with  $\mathbf{y} > \mathbf{0}$ . Note that for a Borel set  $\mathcal{B} \subset \overline{\mathbb{R}}$  we have  $a\mathcal{B} = \{a\mathbf{v} : \mathbf{v} \in \mathcal{B}\}$  and  $\mathcal{B}_{a\mathbf{y}} = a\mathcal{B}_{\mathbf{y}}$ . Now, let

$$\mathbb{W} := (\mathbf{v} \in \overline{\mathbb{R}} : v_1 + \dots + v_d = 1),$$

be the unit simplex on  $\overline{\mathbb{R}}$  (simplex for brevity), where  $d - 1$  variables are free to vary and one is fixed, e.g.  $v_d = 1 - (v_1 + \dots + v_{d-1})$ . For any  $\mathbf{v} \in \mathbb{R}_+^d$ , with the sum-norm,  $\|\mathbf{v}\| = |v_1| + \dots + |v_d|$ , we measure the distance of  $\mathbf{v}$  from  $\mathbf{0}$ . Other norms can also be considered (e.g. Resnick, 2007, pp. 270–274). We consider the one-to-one transformation  $Q : \overline{\mathbb{R}} \rightarrow (0, \infty) \times \mathbb{W}$ , given by

$$(r, \mathbf{w}) := Q(\mathbf{v}) = (\|\mathbf{v}\|, \|\mathbf{v}\|^{-1}\mathbf{v}), \quad \mathbf{v} \in \overline{\mathbb{R}}.$$

By means of this, the induced measure is  $\psi := \eta * Q$ , i.e.  $\psi(\mathcal{W}_r) = \eta\{Q^{\leftarrow}(\mathcal{W}_r)\}$  for all sets

$\mathcal{W}_r = r \times \mathcal{W}$  with  $r > 0$  and  $\mathcal{W} \subset \mathbb{W}$ , is generated. Then, from the property (6) it follows that

$$\begin{aligned}\psi(\mathcal{W}_r) &= \eta\{\mathbf{v} \in \overline{\mathbb{R}} : \|\mathbf{v}\| > r, \mathbf{v}/\|\mathbf{v}\| \in \mathcal{W}\} \\ &= \eta\{r\mathbf{u} \in \overline{\mathbb{R}} : \|\mathbf{u}\| > 1, \mathbf{u}/\|\mathbf{u}\| \in \mathcal{W}\} \\ &= r^{-1}H'(\mathcal{W}),\end{aligned}$$

where  $H'(\mathcal{W}) := \eta\{\mathbf{u} \in \overline{\mathbb{R}} : \|\mathbf{u}\| > 1, \mathbf{u}/\|\mathbf{u}\| \in \mathcal{W}\}$ . The benefit of transforming the coordinates into pseudo-polar is that the measure  $\eta$  becomes a product of two independent measures: the *radial measure* ( $1/r$ ) and *spectral measure* or *angular measure* ( $H'$ ) (e.g. Falk et al., 2011, p. 145). The first measures the intensity (or distance) of the points from the origin and the second measures the angular spread (or direction) of the points. This result is known as the *spectral decomposition* (de Haan and Resnick, 1977). Hereafter we will use the term angular measure.

The density of  $\psi$  is  $d\psi(r, \mathbf{w}) = r^{-2}dr \times dH'(\mathbf{w})$  for all  $r > 0$  and  $\mathbf{w} \in \mathbb{W}$ , by means of which we obtain the explicit form

$$\begin{aligned}\eta(\mathcal{B}_y) &= \psi\{Q(\mathbf{v} \in \overline{\mathbb{R}} : v_1 > y_1 \text{ or } \dots \text{ or } v_d > y_d)\} \\ &= \psi[\{(r, \mathbf{w}) \in (0, \infty) \times \mathbb{W} : r > \min(y_j/w_j, j \in I)\}] \\ &= \int_{\mathbb{W}} \int_{\min(y_j/w_j, j \in I)}^{\infty} r^{-2} dr dH'(\mathbf{w}) \\ &= \int_{\mathbb{W}} \max_{j \in I}(w_j/y_j) dH'(\mathbf{w}).\end{aligned}\tag{7}$$

In pseudo-polar coordinates, extremes are the values whose radial component is higher than a high threshold, see the red points in the middle panels of Figure 1. The angular components are concentrated around the center of the simplex, in the case of strong dependence (middle-top panel), while they are concentrated around the vertices of the simplex (middle-bottom panel), in the case of weak dependence.

The measure  $H'$  can be any finite measure on  $\mathbb{W}$  satisfying the first moment conditions

$$\int_{\mathbb{W}} w_j dH'(\mathbf{w}) = 1, \quad \forall j \in I.$$

This guarantees that the marginal distributions of  $G_0$  are unit Fréchet. If  $H'$  satisfies the first

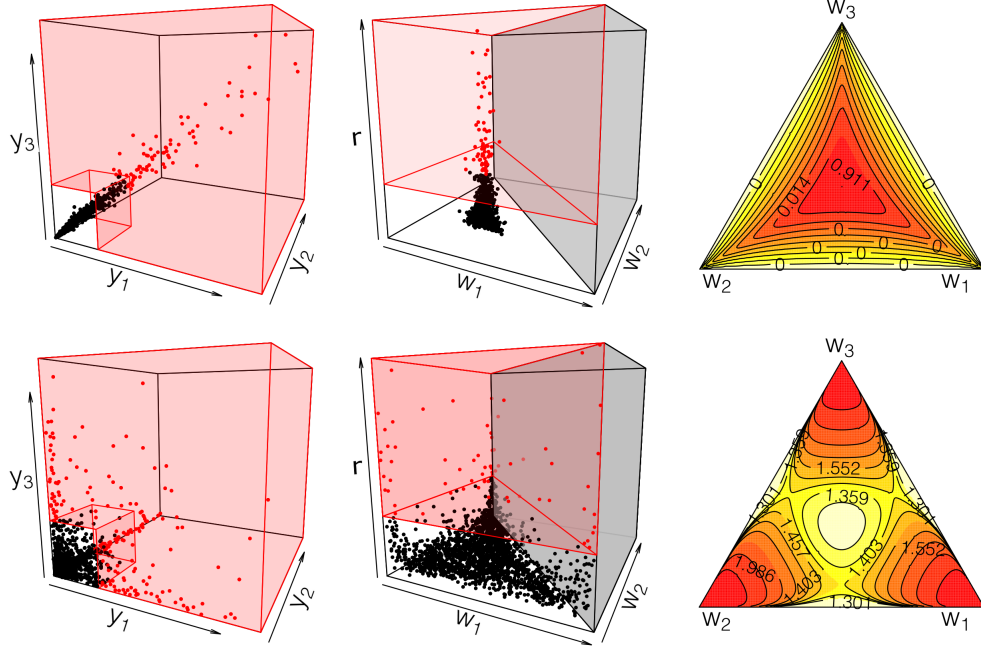


Figure 1: Examples of critical regions in  $\mathbb{R}_+^3$  (left-panels) and its representation in pseudo-polar coordinates (middle-panels). Red points are the extremes with strong (top-panels) and weak (bottom-panels) dependence. Right panels display the angular densities on the simplex.

moment conditions, then the total mass is equal to

$$H'(\mathbb{W}) = \int_{\mathbb{W}} (w_1 + \dots + w_d) dH'(\mathbf{w}) = \sum_{j \in I} \int_{\mathbb{W}} w_j dH'(\mathbf{w}) = d.$$

So setting  $H := H'/H'(\mathbb{W})$ , then  $H$  is a probability measure satisfying

$$\int_{\mathbb{W}} w_j dH(\mathbf{w}) = 1/d, \quad \forall j \in I. \quad (8)$$

Concluding, combining (3), (4), (5) and (7) all together, we have that a MEVD with unit Fréchet margins is equal to

$$G_0(\mathbf{y}) = \exp \left\{ -d \int_{\mathbb{W}} \max_{j \in I} (w_j/y_j) dH(\mathbf{w}) \right\}. \quad (9)$$

## 2.2 Angular densities

The measure  $H$  can place mass on the interior as well as on other subspaces of the simplex, such as the edges and the vertices. Thus  $H$  can have several densities that lie on these sets, which are named *angular densities*. Coles and Tawn (1991) described a way to derive the angular densities



when  $G$  is absolutely continuous (see also Resnick, 2007, Example. 5.13).

Specifically, let  $\mathbb{S} := \mathbb{P}(I) \setminus \emptyset$ , where  $\mathbb{P}(I)$  is the power set of  $I$  and  $\mathcal{S}$  be the index set that takes values in  $\mathbb{S}$ . Given fixed  $d$ , the sets

$$\mathbb{W}_{d,\mathcal{S}} = (\mathbf{w} \in \mathbb{W} : w_j = 0, \text{ if } j \notin \mathcal{S}; w_j > 0 \text{ if } j \in \mathcal{S}),$$

for all  $\mathcal{S} \in \mathbb{S}$  provide a partition of  $\mathbb{W}$  in  $2^d - 1$  subsets. Similar to the simplex, there are  $k - 1$  variables  $w_j$  in  $\mathbb{W}_{d,\mathcal{S}}$  that are free to vary, where  $j \in \mathcal{S}$  and  $k = |\mathcal{S}|$  denotes the size of  $\mathcal{S}$ . We denote by  $h_{d,\mathcal{S}}$  the density that lies on the subspace  $\mathbb{W}_{d,\mathcal{S}}$ , where  $\mathcal{S} \in \mathbb{S}$ . When the latter is a vertex  $\mathbf{e}_j$  of the simplex  $\mathbb{W}$ , for any  $j \in I$ , then the density is a point mass, that is  $h_{d,\mathcal{S}} = H(\{\mathbf{e}_j\})$ .

Let  $\mathcal{S} = \{i_1, \dots, i_k\} \subset I$ , when  $G_0$  is absolutely continuous the angular density for any  $\mathbf{y} \in \mathbb{R}_+^d$  is

$$h_{d,\mathcal{S}} \left( \frac{y_{i_1}}{\sum_{i \in \mathcal{S}} y_i}, \dots, \frac{y_{i_{k-1}}}{\sum_{i \in \mathcal{S}} y_i} \right) = - \left( \sum_{i \in \mathcal{S}} y_i \right)^{(k+1)} \lim_{\substack{y_j \rightarrow 0, \\ j \notin \mathcal{S}}} \frac{\partial^k V}{\partial y_{i_1} \dots \partial y_{i_k}}(\mathbf{y}). \quad (10)$$

Two examples of a tridimensional angular density in the interior of the simplex are reported in the right panels of Figure 1. These are the densities of a symmetric logistic model (Gumbel, 1960) with a strong and weak dependence. When  $\mathcal{S} = \{i\}$  for any  $i \in I$  the angular density  $h_{d,\mathcal{S}}$  represents the mass of  $H$  at the vertex  $\mathbf{e}_j$  with  $j = i$ , thus (10) reduces into

$$h_{d,\mathcal{S}} = H(\{\mathbf{e}_i\}) = -y_i^{(2)} \lim_{y_j \rightarrow 0, j \notin \mathcal{S}} \frac{\partial V}{\partial y_i}(\mathbf{y}). \quad (11)$$

In the bivariate case these results are equal to the ones obtained by Pickands (1981). Kotz and Nadarajah (2000) discussed the bivariate case in the following terms. With  $d = 2$  the unit simplex  $\mathbb{W} = [0, 1]$  can be partitioned into

$$\mathbb{W}_{2,\{1\}} = \{(1, 0)\}, \quad \mathbb{W}_{2,\{2\}} = \{(0, 1)\}, \quad \mathbb{W}_{2,\{1,2\}} = \{(w, 1 - w), w \in (0, 1)\}.$$

The densities that lie on them are

$$h_{2,\{1\}} = H(\{0\}) = -y_1^2 \lim_{y_2 \rightarrow 0} \frac{\partial V}{\partial y_1}(y_1, y_2),$$

$$h_{2,\{2\}} = H(\{1\}) = -y_2^2 \lim_{y_1 \rightarrow 0} \frac{\partial V}{\partial y_2}(y_1, y_2),$$

and

$$h_{2,\{1,2\}}(w) = -\frac{\partial^2 V}{\partial y_1 \partial y_2}(w, 1-w).$$

respectively, for any  $y_1, y_2 > 0$ . The first two densities describe the case when extremes are only observed in one variable. While the third density describes the case when extremes are observed in both variables.

### 2.3 Extremal dependence

From (5) it emerges that the extremal dependence is expressed through the exponent function. This is a map from  $\mathbb{R}_+^d$  to  $(0, \infty)$  satisfying the properties:

1. is a continuous function and homogeneous of order  $-1$ , the latter meaning that  $V(a\mathbf{y}) = a^{-1}V(\mathbf{y})$  for all  $a > 0$ ;
2. is a convex function, that is  $V(a\mathbf{y} + (1-a)\mathbf{y}') \leq aV(\mathbf{y}) + (1-a)V(\mathbf{y}')$ , for  $a \in [0, 1]$  and  $\mathbf{y}, \mathbf{y}' \in \mathbb{R}_+^d$ ;
3.  $\max(1/y_1, \dots, 1/y_d) \leq V(\mathbf{y}) \leq (1/y_1 + \dots + 1/y_d)$ , with the lower and upper limits representing the complete dependence and independence cases respectively.

See de Haan and Ferreira (2006, pp. 223–226) for details. In summary, let  $\mathbf{Y}$  be a random vector with distribution (9). When  $H$  places the total mass 1 on the center of the simplex  $(1/d, \dots, 1/d)$ , then  $Y_1 = Y_2 = \dots = Y_d$  almost surely and hence  $G_0(\mathbf{y}) = \exp\{\max(1/y_1, \dots, 1/y_d)\}$ . When  $H$  places mass  $1/d$  on  $\mathbf{e}_j$  for all  $j \in I$ , i.e. the vertices of the simplex, then  $Y_1, \dots, Y_d$  are independent and hence  $G_0(\mathbf{y}) = \exp(1/y_1 + \dots + 1/y_d)$ . This rephrased for a random vector  $\mathbf{X}$  with distribution (1) becomes

$$\min\{G_1(x_1), \dots, G_d(x_d)\} \leq G(\mathbf{x}) \leq G_1(x_1) \cdot \dots \cdot G_d(x_d), \quad \mathbf{x} \in \mathbb{R}^d.$$

In order to visualise the exponent function more easily, its restriction in the simplex is usually considered. This is a function  $A : \mathbb{W} \rightarrow [1/d, 1]$ , named the *Pickands dependence* function (Pickands, 1981), defined by

$$A(\mathbf{t}) := d \int_{\mathbb{W}} \max_{j \in I} (w_j t_j) dH(\mathbf{w}),$$

where  $z_j = 1/y_j$ ,  $j \in I$ ,  $t_j = z_j/(z_1 + \dots + z_d)$  with  $j = 1, \dots, d-1$  and  $t_d = 1 - (t_1 + \dots + t_{d-1})$ .  $A$  inherits the above properties from  $V$  with the obvious modifications. In particular,  $1/d \leq$

$\max(t_1, \dots, t_d) \leq A(\mathbf{t}) \leq 1$ , where lower and upper bounds represent the complete dependence and independence cases, and for the homogeneity property of  $A$  the exponent function can be rewritten as

$$V(\mathbf{z}) = (z_1 + \dots + z_d)A(t_1, \dots, t_d), \quad \mathbf{z} \in \mathbb{R}_+^d.$$

The exponential function can be profitably used in several ways. First, an important summary of the extremal dependence is given by

$$\vartheta = V(1, \dots, 1) = d \int_{\mathbb{W}} \max_{j \in I}(w_j) dH(\mathbf{w}). \quad (12)$$

This is named the *extremal coefficient* (Smith, 1990) and it represents the (fractional) number of independent components of the random vector  $\mathbf{Y}$ . The coefficient takes values in  $[1, d]$ , depending on whether the measure  $H$  concentrates near the center or the vertices of the simplex. The bounds regard the cases of complete dependence and independence.

Second, for any  $\mathbf{y} > \mathbf{0}$  and failure region

$$\mathcal{F}_{\mathbf{y}} = (\mathbf{v} \in \overline{\mathbb{R}} : v_1 > y_1 \text{ and } \dots \text{ and } v_d > y_d), \quad (13)$$

the *tail dependence* function (Nikoloulopoulos et al., 2009; de Haan and Ferreira, 2006, p. 225) is defined by

$$R(\mathbf{y}) := \eta\{(\mathbf{v} \in \overline{\mathbb{R}} : v_1 > y_1 \text{ and } \dots \text{ and } v_d > y_d)\} \equiv \eta(\mathcal{F}_{\mathbf{y}}), \quad \mathbf{y} > \mathbf{0}.$$

This counts the number of observations that fall in the failure region, i.e. all their coordinates are simultaneously large. The tail dependence function is related to the exponent function by the inclusion-exclusion principle. Using similar arguments to those in (7) and (8) it follows that

$$R(\mathbf{y}) = d \int_{\mathbb{W}} \min_{j \in I}(w_j/y_j) dH(\mathbf{w}) \quad \mathbf{y} > \mathbf{0}. \quad (14)$$

By means of the tail dependence function, another important summary of the dependence between the components of  $\mathbf{Y}$  is obtained. The *coefficient of upper tail dependence* is given by

$$\chi = R(1, \dots, 1) = d \int_{\mathbb{W}} \min_{j \in I}(w_j) dH(\mathbf{w}). \quad (15)$$

It measures the strength of dependence in the tail of the distribution of  $\mathbf{Y}$  or in other terms the probability that all the components of  $\mathbf{Y}$  are simultaneously large. This coefficient was introduced in the bivariate case by Joe (1997, Ch. 2) and extended to the multivariate case by Li (2009). When  $H$  concentrates near the center or on the vertices of the simplex, then  $\chi > 0$  or  $\chi = 0$  respectively. In these cases we say that  $\mathbf{Y}$  is upper tail dependent or independent.

In addition, the exponent and the tail dependence functions can be used for approximating the probability that certain types of extreme events will occur. Specifically, let  $\mathbf{Y}$  be a random vector with unit Pareto margins.  $F$  is in the domain of attraction of a MEVD with Fréchet margins. From (4) and for the homogeneity property of  $V$  we have that  $\{1 - F(n\mathbf{y})\} \approx V(n\mathbf{y})$  for large  $n$ . Then, for the relations (7) and (8), the approximating result follows

$$\text{pr}(Y_1 > y_1 \text{ or } \dots \text{ or } Y_d > y_d) \approx d \int_{\mathbb{W}} \max_{j \in I} (w_j / y_j) dH(\mathbf{w}), \quad (16)$$

when  $y_1, \dots, y_d$  are high enough thresholds. Furthermore, with similar arguments to those in Section 2.1 we have that

$$\lim_{n \rightarrow \infty} n\bar{F}(ny_1, \dots, ny_d) = R(\mathbf{y}),$$

where  $\bar{F}$  is the survivor function of  $\mathbf{Y}$ .  $R$  has the same homogeneity property of  $V$ . Hence,  $\bar{F}(n\mathbf{y}) \approx R(n\mathbf{y})$  for large  $n$ . Then, for the relation (14), the approximating result also follows

$$\text{pr}(Y_1 > y_1 \text{ and } \dots \text{ and } Y_d > y_d) \approx d \int_{\mathbb{W}} \min_{j \in I} (w_j / y_j) dH(\mathbf{w}), \quad (17)$$

when  $y_1, \dots, y_d$  are high enough thresholds.

Lastly, when  $\chi = 0$  the elements of  $\mathbf{Y}$  are independent in the limit. However, they may still be dependent for large but finite samples. Ledford and Tawn (1996) proposed another dependence measure in order to capture this feature. For brevity, we focus on the bivariate case. Suppose that  $\bar{F}$  for  $y \rightarrow \infty$  satisfies the condition

$$\bar{F}(y, y) \approx y^{-1/\tau} \mathcal{L}(y), \quad 0 < \tau \leq 1,$$

where  $\mathcal{L}$  is a slowly function, i.e.  $\mathcal{L}(ay)/\mathcal{L}(y) \rightarrow 1$  as  $y \rightarrow \infty$  for any  $a > 0$ . Then for large  $y$ , assuming  $\mathcal{L}$  constant, different tail behaviours are covered. The case  $\chi > 0$  is reached when  $\tau = 1$  and so the variables are asymptotically dependent. When  $1/2 < \tau < 1$  this means that  $\chi = 0$  and so the variables are asymptotically independent, but they are still positively associated and the

value of  $\tau$  expresses the degree (see Ledford and Tawn, 1996, for details).

### 3 Parametric models for the extremal dependence

From the previous sections, it emerges that both the exponent and tail dependence functions depend on the angular measure. There is no unique angular measure that generates the extremal dependence, any finite measure that satisfies the first moment conditions is suitable. In order to represent the extremal dependence, in principle it is insufficient to use a parametric family of models for the distribution function of the angular measure. However, flexible classes of parametric models can still be useful for applications, e.g. see Tawn (1990), Coles and Tawn (1991) and Boldi and Davison (2007) to name a few. To this end, in previous years different parametric extremal dependence models have been introduced in the literature. A fairly comprehensive overview can be found in Kotz and Nadarajah (2000, Section 3.4), Coles (2001, Section 8.2.1), Beirlant et al. (2006, Section 9.2.2) and Padoan (2013b). In the next sections we describe some of the most popular models.

#### 3.1 Asymmetric logistic model

The multivariate asymmetric logistic model is an extension of the symmetric, introduced by Tawn (1990) (see also Coles and Tawn, 1991) for modelling extremes in complex environmental applications.

Let  $\mathbb{S}$  and  $\mathcal{S}$  as in Section 2.2 and  $N_{\mathcal{S}}$  be a Poisson random variable with rate  $1/\tau_{\mathcal{S}}$ . This describes the number of storm events,  $n_{\mathcal{S}}$ , that takes place on the sites  $\mathcal{S}$  in a time interval. Given  $n_{\mathcal{S}}$ , for any site  $j \in \mathcal{S}$ , let  $\{X_{j,\mathcal{S};i}, i = 1, \dots, n_{\mathcal{S}}\}$  be a sequence of i.i.d. random variables that describe an environmental episode such as rain. For a fixed  $i$ ,  $\{X_{j,\mathcal{S};i}\}_{j \in \mathcal{S}}$  is assumed to be a dependent sequence. The maximum amount of rain observed at  $j$  is  $X_{j,\mathcal{S}} = \max_{i=1, \dots, n_{\mathcal{S}}} \{X_{j,\mathcal{S};i}\}$ . Let  $A_{\mathcal{S}}$  be a random effect with a positive stable distribution and stability parameter  $\alpha_{\mathcal{S}} \geq 1$  (Nolan, 2003), representing an unrecorded additional piece of information on storm events. Assume  $\{X_{j,\mathcal{S}}\}_{j \in \mathcal{S}} | \alpha_{\mathcal{S}}$  as an independent sequence. Define  $Y_j = \max_{\mathbb{S} \in \mathbb{S}_j} \{X_{j,\mathcal{S}}\}$ , where  $\mathbb{S}_j \subset \mathbb{S}$  contains all nonempty sets including  $j$  and so the maximum is over all the storm events involving  $j$ . Then, the exponent function of the joint survival function of  $(Y_1, \dots, Y_d)$ , after transforming the margins into unit

exponential variables, is

$$V(\mathbf{y}; \boldsymbol{\theta}) = \sum_{\mathcal{S} \in \mathbb{S}} \left\{ \sum_{j \in \mathcal{S}} (\beta_{j,\mathcal{S}} y_j^{-1})^{\alpha_{\mathcal{S}}} \right\}^{1/\alpha_{\mathcal{S}}}, \quad \mathbf{y} \in \mathbb{R}_+^d,$$

where  $\boldsymbol{\theta} = \{\alpha_{\mathcal{S}}, \beta_{j,\mathcal{S}}\}_{\mathcal{S} \in \mathbb{S}}$ ,  $\alpha_{\mathcal{S}} \geq 1$ ,  $\beta_{\mathcal{S}} = \tau_{\mathcal{S}} / \sum_{\mathcal{S} \in \mathbb{S}_j} \tau_{\mathcal{S}}$  and  $\beta_{j,\mathcal{S}} = 0$  if  $j \notin \mathcal{S}$ , and for  $j \in \mathcal{S}$ ,  $0 \leq \beta_{j,\mathcal{S}} \leq 1$  and  $\sum_{\mathcal{S} \in \mathbb{S}} \beta_{j,\mathcal{S}} = 1$ . The parameter  $\beta_{j,\mathcal{S}}$  represents the probability that the maximum value observed at  $j$  is attributed to a storm event involving the sites of  $\mathcal{S}$ . The number of the model parameters is  $2^{d-1}(d+2) - (2d+1)$ .

In this case the angular measure places mass on all the subspaces of the simplex. From (10) it follows that the angular density is, for every  $\mathcal{S} \in \mathbb{S}$  and all  $\mathbf{w} \in \mathbb{W}_{d,\mathcal{S}}$  equal to

$$h_{d,\mathcal{S}}(\mathbf{w}; \boldsymbol{\theta}) = \prod_{i=1}^{k-1} (i\alpha_{\mathcal{S}} - 1) \prod_{j \in \mathcal{S}} \beta_{j,\mathcal{S}}^{\alpha_{\mathcal{S}}} w_j^{-(\alpha_{\mathcal{S}}+1)} \left\{ \sum_{j \in \mathcal{S}} (\beta_{j,\mathcal{S}}/w_j)^{\alpha_{\mathcal{S}}} \right\}^{1/\alpha_{\mathcal{S}}-k}.$$

When  $\mathcal{S} = I$ ,  $\alpha_{\mathcal{S}} = \alpha$ ,  $\beta_{j,\mathcal{S}} = \beta_j$  and so the angular density on the interior of the simplex simplifies to

$$h(\mathbf{w}; \boldsymbol{\theta}) = \prod_{i=1}^{d-1} (i\alpha - 1) \prod_{j \in I} \beta_j^{\alpha} w_j^{-(\alpha+1)} \left\{ \sum_{j \in I} (\beta_j/w_j)^{\alpha} \right\}^{1/\alpha-d}, \quad \mathbf{w} \in \mathbb{W}.$$

When  $\mathcal{S} = \{j\}$ , for all  $j \in I$ , then from (11) it follows that the point mass at each extreme point of the simplex is  $h_{d,\mathcal{S}} = \beta_{j,s}$ .

For example in the bivariate case, the conditions on the parameters are  $\beta_{1,\{1\}} + \beta_{1,\{1,2\}} = 1$  and  $\beta_{2,\{2\}} + \beta_{2,\{1,2\}} = 1$ , so the masses at the corners of  $\mathbb{S}_2 = [0, 1]$  are given by  $h_{2,\{1\}} = 1 - \beta_1$  and  $h_{2,\{2\}} = 1 - \beta_2$ , where for simplicity  $\beta_{1,\{1,2\}} = \beta_1$  and  $\beta_{2,\{1,2\}} = \beta_2$ , while the density in the interior of the simplex, for  $0 < w < 1$ , is

$$h_{2,\{1,2\}}(w) = (\alpha - 1)(\beta_1 \beta_2)^{\alpha} \{w(1-w)\}^{\alpha-2} [(\beta_1(1-w))^{\alpha} + (\beta_2 w)^{\alpha}]^{1/\alpha-2}.$$

The top row of Figure 2 illustrates some examples of trivariate angular densities for different values of the parameters  $\boldsymbol{\theta} = (\alpha, \beta_1, \beta_2, \beta_3)$ , where the subscript of the index set  $\mathcal{S} = \{1, 2, 3\}$  has been omitted for simplicity. The values of the parameters are, from left to right  $\{(5.75, 0.5, 0.5, 0.5); (1.01, 0.9, 0.9, 0.9); (1.25, 0.5, 0.5, 0.5); (1.4, 0.7, 0.15, 0.15)\}$ . The first panel shows that with large values of  $\alpha$  and equal values of the other parameters, the case of strong dependence among the variables is obtained. The mass is mainly concentrated towards the center of the simplex. The second panel shows that when  $\alpha$  is close to 1 and the other parameters are equal,

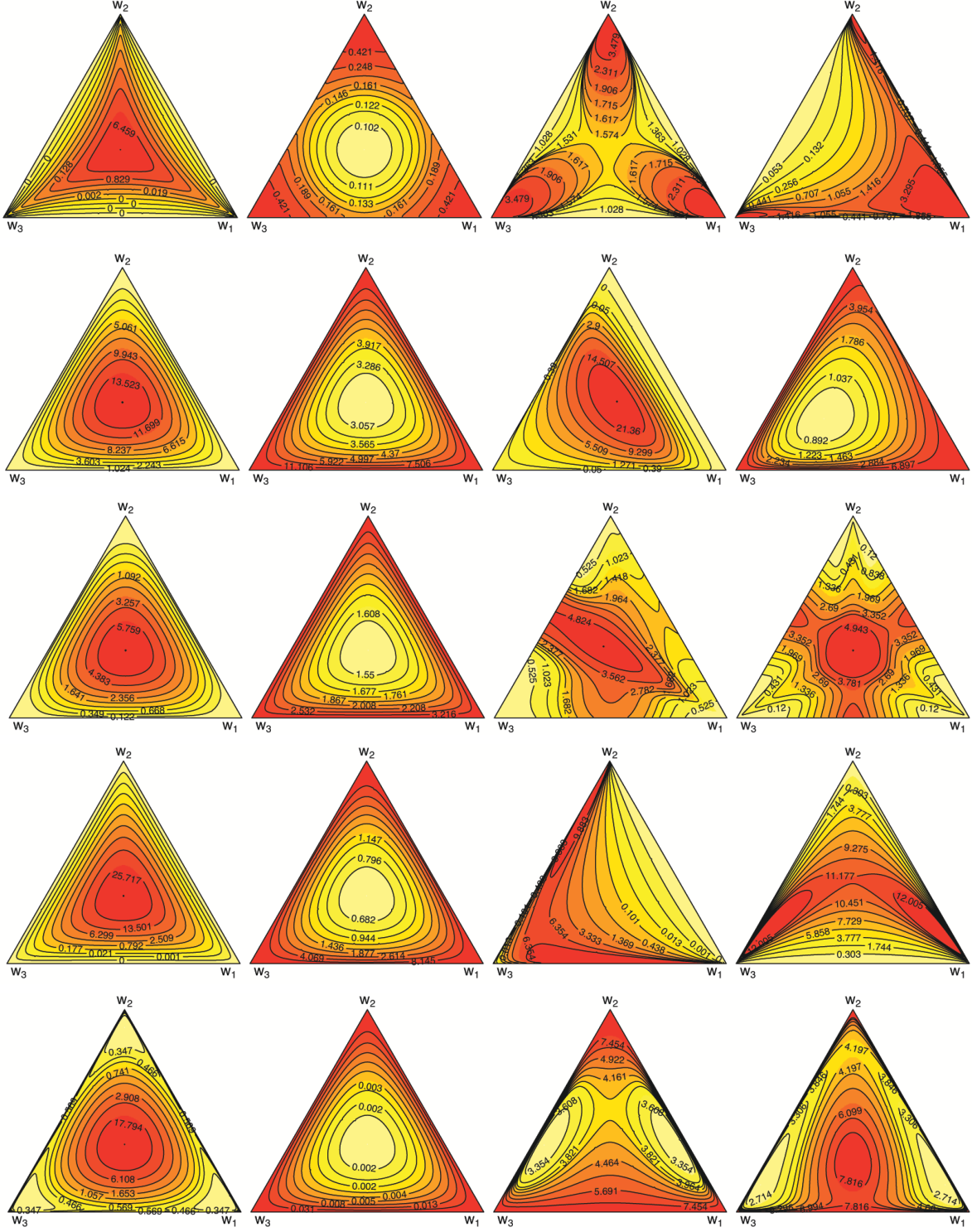


Figure 2: Examples of trivariate angular densities for the Asymmetric Logistic, Tilted Dirichlet, Pairwise Beta, Hüsler-Reiss and Extremal- $t$  models from top to bottom.

the case of weak dependence is obtained. The mass is concentrated on the vertices of the simplex. The third panel shows the case of a symmetric dependence structure with the mass near the corners of the simplex but not along the edges. Finally, the fourth panel shows a case of an asymmetric dependence structure where the mass tends to be closer to the components whose corresponding values of  $\beta$  are high.

### 3.2 Tilted Dirichlet model

Extremal dependence models with an angular measure that places mass on the interior, vertices and edges of the simplex are more flexible than those with a measure that concentrates only on the interior. An example is the asymmetric logistic model versus the symmetric. However, the former has too many parameters to estimate, so parsimonious models may be preferred. In order to derive a parametric model for the angular density whose mass concentrates on the interior of the simplex, Coles and Tawn (1991) proposed the following method. Consider a continuous function  $h' : \mathbb{W} \rightarrow [0, \infty)$  such that  $m_j = \int_{\mathbb{S}_d} v_j h'(\mathbf{v}) d\mathbf{v} < \infty$  for all  $j \in I$ . Then, the function

$$h(\mathbf{w}) = d^{-1} (m_1 w_1 + \dots + m_d w_d)^{-(d+1)} h' \{ \mathbf{m} \mathbf{w} / (m_1 w_1 + \dots + m_d w_d) \}, \mathbf{w} \in \mathbb{W}$$

is a valid angular density. It satisfies the first moment conditions (8) and its mass is centered at  $(1/d, \dots, 1/d)$  and integrates to one. For example, if  $h'$  is the density of the Dirichlet distribution, then we obtain the angular density

$$h(\mathbf{w}; \boldsymbol{\theta}) = \frac{\Gamma(\sum_{j \in I} \alpha_j + 1)}{d(\sum_{j \in I} \alpha_j w_j)^{d+1}} \prod_{j=1}^d \frac{\alpha_j}{\Gamma(\alpha_j)} \left( \frac{\alpha_j w_j}{\sum_{j \in I} \alpha_j w_j} \right)^{\alpha_j - 1}, \mathbf{w} \in \mathbb{W}, \quad (18)$$

where  $\boldsymbol{\theta} = \{\alpha_j > 0\}_{j \in I}$ . This density is asymmetric and it becomes symmetric when  $\alpha_1 = \dots = \alpha_d$ . Extremes are independent or completely dependent when for all  $j \in I$  the limiting cases  $\alpha_j \rightarrow 0$  and  $\alpha_j \rightarrow \infty$  arise. The dependence parameters  $\alpha_j$ ,  $j \in I$ , are not easy to interpret. However, Coles and Tawn (1994) draw attention to the quantities  $r_1 = (\alpha_i - \alpha_j)/2$  and  $r_2 = (\alpha_i + \alpha_j)/2$  which can be interpreted as the asymmetry and intensity of the dependence between pairs of variables.

In this case, the exponent function can not be analytically computed, nonetheless it can still be evaluated numerically.

The second row of Figure 2 illustrates some examples of trivariate angular densities obtained with different sets of the parameters  $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3)$ . The plots from left to right have been obtained



using the parameter sets  $\{(2, 2, 2); (0.5, 0.5, 0.5); (2, 2.5, 30); (0.1, 0.25, 0.95)\}$ . The first panel shows that, when values of the parameters are equal and greater than 1, the mass concentrates in the center of the simplex leading to strong dependence. The second panel shows the opposite, when values of  $\alpha$  are equal and less than 1, it yields to the case of weak dependence as the mass concentrates on the vertices of the simplex. The third panel shows the case of an asymmetric dependence structure and this is obtained when the values of the parameters are all greater than one. In this specific case the mass tends to spread towards the bottom and top left edges. The fourth panel illustrates another case of an asymmetric dependence structure, in this case obtained with all the values of the parameters that are less than 1, leading to a mass that concentrates along the top right edge and vertices.

### 3.3 Pairwise beta model

The tilted Dirichlet model has been successfully used for applications (e.g. Coles and Tawn, 1991), although it suffers from a lack of interpretability of the parameters. Cooley et al. (2010) proposed a similar model but with easily interpretable parameters. The definition of their model is based on a geometric approach. Specifically, they considered the symmetric pairwise beta function

$$h^*(w_i, w_j) = \frac{\Gamma(2\beta_{i,j})}{\Gamma^2(\beta_{i,j})} \left( \frac{w_i}{w_i + w_j} \right)^{\beta_{i,j}-1} \left( \frac{w_j}{w_i + w_j} \right)^{\beta_{i,j}-1}, i, j \in I,$$

where  $w_i$  and  $w_j$  are two elements of  $\mathbf{w}$  and  $\beta_{i,j} > 0$ . This function has its center at the point  $(1/d, \dots, 1/d)$  and it verifies the first moment conditions (8). Then, the angular pairwise beta density is defined by summing together all the  $d(d-1)/2$  possible pairs of variables, namely

$$h(\mathbf{w}; \boldsymbol{\theta}) = \frac{2(d-3)! \Gamma(\alpha d + 1)}{d(d-1) \Gamma(2\alpha + 1) \Gamma\{\alpha(d-2)\}} \sum_{i,j \in I, i < j} h(w_i, w_j), \quad \mathbf{w} \in \mathbb{W},$$

where

$$h(w_i, w_j) = (w_i + w_j)^{2\alpha-1} \{1 - (w_i + w_j)\}^{\alpha(d-2)-d+2} h^*(w_i, w_j)$$

and  $\boldsymbol{\theta} = (\alpha, \{\beta_{i,j}\}_{i,j \in I})$  with  $\alpha > 0$ . Each parameter  $\beta_{i,j}$  controls the level of dependence between the  $i^{th}$  and the  $j^{th}$  components and the dependence increases for increasing values of  $\beta_{i,j}$ . The function  $h^*$  is introduced to guarantee that the dependence ranges between weak and strong dependence. The parameter  $\alpha$  controls the dependence of all the variables, when it increases the overall dependence increases.

Also in this case the exponent function can not be computed in closed form and hence it can only be evaluated numerically.

The third row of Figure 2 provides some examples of trivariate angular densities obtained with different values of the parameters  $\boldsymbol{\theta} = (\alpha, \beta_{1,2}, \beta_{1,3}, \beta_{2,3})$ . The plots from left to right have been obtained using the parameter sets  $\{(4, 2, 2, 2); (0.5, 1, 1, 1); (1, 2, 4, 15); (1, 10, 10, 10)\}$ . The first panel shows a case of symmetric density obtained with all equal parameters  $\beta_{i,j}$   $i, j \in I$ . A large value of the overall dependence parameter  $\alpha$  pulls the mass towards the center of the simplex, indicating a strong dependence between the variables. On the contrary, the second panel shows that when the overall dependence parameter is close to zero then the mass concentrates on the vertices of the simplex, indicating weak dependence among the variables. The third panel illustrates a case of asymmetric angular density with strong dependence between the second and third variables that is due to a large value of  $\beta_{2,3}$ . Although the value of the global dependence parameter  $\alpha$  is not large, it is enough to slightly push the mass towards the center of the simplex. The fourth panel shows a case of symmetric angular density, which is obtained with large values of the pairwise dependence parameters and an average value of the global dependence parameter. The mass is mainly concentrated on the center of the simplex and some mass tends to lie near the centers of the edges.

### 3.4 Hüsler-Reiss model

One of the most popular models is the Hüsler-Reiss (Hüsler and Reiss, 1989). Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be  $n$  i.i.d. copies of a zero-mean unit variance Gaussian random vector. Assume that for all  $i, j \in I$  the pairwise correlation  $\rho_{i,j;n}$  satisfies the condition

$$\lim_{n \rightarrow \infty} \log n(1 - \rho_{i,j;n}) = \lambda_{i,j}^2 \in [0, \infty).$$

Then, the exponent function of the limit distribution of  $\mathbf{b}_n(\mathbf{M}_n - \mathbf{b}_n)$  for  $n \rightarrow \infty$ , where  $\mathbf{b}_n = (b_n, \dots, b_n)$  is a vector of real sequences (see Resnick, 2007, pp. 71-72), is

$$V(\mathbf{y}; \boldsymbol{\theta}) = \sum_{j=1}^d \frac{1}{y_j} \Phi_{d-1} \left\{ \left( \lambda_{i,j} + \frac{\log y_i / y_j}{2\lambda_{i,j}} \right)_{i \in I_j}; \bar{\Lambda}_j \right\}, \quad \mathbf{y} \in \mathbb{R}_+^d, \quad (19)$$

where  $\boldsymbol{\theta} = \{\lambda_{i,j}\}_{i,j \in I}$ ,  $I_j := I \setminus \{j\}$ ,  $\Phi_{d-1}$  is  $d-1$  dimensional Gaussian distribution with partial correlation  $\bar{\Lambda}_j$ . For all  $j \in I$ , the elements of  $\bar{\Lambda}_j$  are  $\lambda_{k,i;j} = (\lambda_{k,j}^2 + \lambda_{i,j}^2 - \lambda_{k,i}^2) / (2\lambda_{k,j}\lambda_{i,j})$ , for

$k, i \in I_j$ . The parameter  $\lambda_{i,j}$ ,  $i, j \in I$ , controls the dependence between the  $i^{th}$  and  $j^{th}$  elements of a vector of  $d$  extremes. These are completely dependent when  $\lambda_{ij} = 0$  and become independent as  $\lambda_{ij} \rightarrow \infty$ .

In this case the angular measure concentrates on the interior of the simplex. Applying (10) it can be checked (Engelke et al., 2015) that the angular density is

$$h(\mathbf{w}; \boldsymbol{\theta}) = \phi_{d-1} \left\{ \left( \lambda_{i,1} + \frac{\log w_i/w_1}{2\lambda_{i,1}} \right)_{i \in I_1}; \bar{\Lambda}_1 \right\} \left\{ w_1^2 \prod_{i=2}^d (w_i 2\lambda_{i,1}) \right\}^{-1}, \mathbf{w} \in \mathbb{W},$$

where  $\phi_{d-1}$  is  $d - 1$  dimensional Gaussian density with partial correlation matrix  $\bar{\Lambda}_1$ .

The second last row of Figure 2 provides some examples of trivariate angular densities obtained with different values of the parameters  $\boldsymbol{\theta} = (\lambda_{1,2}, \lambda_{1,3}, \lambda_{2,3})$ . The plots from left to right have been obtained using the parameter sets  $\{(0.3, 0.3, 0.3), (1.4, 1.4, 1.4), (1.7, 0.7, 1.1), (0.52, 0.71, 0.52)\}$ . The first panel shows that with small and equal values of parameters the case of strong dependence among all the variables is obtained. In this case the mass concentrates around the center of the simplex. On the contrary, the second panel shows that with large and equal values of the parameters the case of weak dependence is obtained. In this case the mass is placed close to the vertices of the simplex. The third panel shows that an asymmetric dependence structure is obtained when the parameter values are different. In this case the mass tends to concentrate around the vertices and edges that are concerned with the smaller values of the parameters. The fourth panel shows that a symmetric dependence structure, with respect to the second component is obtained setting the values of two parameters to be equal. In this case the mass is equally divided up towards the two vertices and edges that are concerned with the smaller values of the parameters.

### 3.5 Extremal- $t$ model

The extremal- $t$  model (Nikoloulopoulos et al., 2009) is more flexible than the Hüsler-Reiss but it is still simple enough. It is easily interpretable and useful in practical applications (see Davison et al., 2012). Let  $\mathbf{X}_1, \dots, \mathbf{X}_n$  be  $n$  i.i.d. copies of a zero-center unit scale Student- $t$  random vector with dispersion matrix  $\Sigma$  and  $\nu > 0$  degrees of freedom (d.f.). Then, the exponent function of the limiting distribution of  $\mathbf{M}_n/\mathbf{a}_n$  for  $n \rightarrow \infty$ , where  $\mathbf{a}_n = (a_n, \dots, a_n)$  is a vector of positive sequences

(see Demarta and McNeil, 2005), is

$$V(\mathbf{y}; \boldsymbol{\theta}) = \sum_{j=1}^d \frac{1}{y_j} T_{d-1, \nu+1} \left\{ \left[ \sqrt{\frac{\nu+1}{1-\rho_{i,j}^2}} \left\{ \left( \frac{y_i}{y_j} \right)^{\frac{1}{\nu}} - \rho_{i,j} \right\} \right]_{i \in I_j}; \bar{\Sigma}_j \right\}, \quad (20)$$

for all  $\mathbf{y} \in \mathbb{R}_+^d$ , where  $\boldsymbol{\theta} = (\{\rho_{i,j}\}_{i,j \in I}, \nu)$  and  $T_{d-1, \nu+1}$  is a  $d-1$  dimensional Student- $t$  distribution with  $\nu+1$  d.f. and partial correlation matrix  $\bar{\Sigma}_j$ . The correlation parameter  $\rho_{i,j}$ ,  $i, j \in I$ , drives the dependence between pairs of variables with the dependence that increases with the increasing of  $\rho_{i,j}$ . The parameter  $\nu$  controls the overall dependence among all the variables. For decreasing values of  $\nu$  the dependence increases and vice versa.

The Hüsler-Reiss model is a special case of the extremal- $t$ . Indeed, for all  $i, j \in I$  if the correlation parameters of the extremal- $t$  distribution are equal to  $\rho_{i,j;\nu} = 1 - \lambda_{i,j}^2/\nu$ , then this distribution converges weakly, as  $\nu \rightarrow \infty$ , to the Hüsler-Reiss (see Nikoloulopoulos et al., 2009).

In this case the angular measure places mass on all the subspaces of the simplex. When  $\mathcal{S} = I$ , then applying (10) we obtain that the angular density is

$$h(\mathbf{w}; \boldsymbol{\theta}) = \frac{t_{d-1, \nu+1} \left( \left[ \sqrt{\frac{\nu+1}{1-\rho_{i,1}^2}} \left\{ (w_i/w_1)^{1/\nu} - \rho_{i,1} \right\} \right]_{i \in I_1}; \bar{\Sigma}_1 \right)}{\nu^{d-1} w_1^{d+1} \left\{ \prod_{i=2}^d \sqrt{\frac{\nu+1}{1-\rho_{i,1}^2}} (w_i/w_1)^{(\nu-1)/\nu} \right\}^{-1}}, \quad \mathbf{w} \in \mathbb{W},$$

where  $t_{d-1, \nu+1}$  is  $d-1$  dimensional Student- $t$  density with partial correlation matrix  $\bar{\Sigma}_1$  (e.g. Ribatet, 2013). When  $\mathcal{S} = \{j\}$ , then applying (11) we obtain that the mass on the extreme points of the simplex is

$$h_{d,\mathcal{S}} = T_{d-1, \nu+1} \left[ \left\{ -\rho_{i,j}(\nu+1)^{1/2}/(1-\rho_{i,j}^2)^{1/2} \right\}_{i \in I_j}; \bar{\Sigma}_j \right], \quad j \in I.$$

The last row of Figure 2 provides some examples of the trivariate angular densities obtained with different values of the parameters  $\boldsymbol{\theta} = (\rho_{1,2}, \rho_{1,3}, \rho_{2,3}, \nu)$ . From left to right the plots are obtained using the parameter values  $\{(0.95, 0.95, 0.95, 2); (-0.3, -0.3, -0.3, 5); (0.52, 0.71, 0.52, 3); (0.52, 0.71, 0.52, 2)\}$ . The first panel shows that when the scale parameters  $\rho_{ij}$  are all equal and close to one and the d.f.  $\nu$  are small, then the mass concentrates around the center of the simplex and therefore the dependence is strong. The second panel shows the opposite, when the correlations are close to zero and the d.f. are high, the mass concentrates around the vertices of the simplex and hence the dependence is weak. The third panel shows that when two scale parameters are equal then the dependence structure is symmetric with respect to the second component and the

mass tends to concentrate on the top vertex and the bottom edge and vertices. The fourth panel shows that with the same setting but with smaller d.f. the mass is pushed towards the center of the simplex and hence the dependence is stronger.

## 4 Estimating the extremal dependence

Several inferential methods have been explored for inferring the extremal dependence. Nonparametric and parametric approaches are available. In the first case recent advances are Gudendorf and Segers (2011), Gudendorf and Segers (2012) and Marcon et al. (2014), see also the references therein. Both likelihood based and Bayesian inferential methods have been widely investigated. Examples of likelihood based methods are the approximate likelihood (e.g. Coles and Tawn, 1994; Cooley et al., 2010; Engelke et al., 2015) and the composite likelihood (e.g. Padoan et al., 2010; Davison and Gholamrezaee, 2012). Examples of Bayesian techniques are Apputhurai and Stephenson (2011), Sabourin et al. (2013), Sabourin and Naveau (2014).

For comparison purposes in the next section the real data analysis is performed using the maximum approximate likelihood estimation method and the approximate Bayesian method based on the approximate likelihood. Here is a brief description.

From the theory in Sections 2.1, if  $\mathbf{Y}_1, \dots, \mathbf{Y}_n$  are i.i.d. copies of  $\mathbf{Y}$  on  $\mathbb{R}_+^d$  with a distribution in the domain of attraction of a MEVD, then the distribution of the sequence  $\{R_i/n, \mathbf{W}_i, i = 1, \dots, n\}$ , where  $R_i = Y_{i,1} + \dots + Y_{i,d}$  and  $\mathbf{W}_i = \mathbf{Y}_i/R_i$ , converges as  $n \rightarrow \infty$  to the distribution of a PPP with density  $d\psi(r, \mathbf{w}) = r^{-2}dr \times dH(\mathbf{w})$ .

Assume that  $\mathbf{x}_1, \dots, \mathbf{x}_n$  are i.i.d. observations from a random vector with an unknown distribution. Since the aim is estimating the extremal dependence, we transform the data into the sample  $\mathbf{y}_1, \dots, \mathbf{y}_n$  with unit Fréchet marginal distributions. This is done by applying the probability integral transform, after fitting the marginal distributions. Next, the coordinates of the data-points are changed from Euclidean into pseudo-polar by the transformation

$$r_i = y_{i,1} + \dots + y_{i,d} \quad \mathbf{w}_i = \mathbf{y}_i/r_i, \quad i = 1, \dots, n.$$

Then, the sequence  $\{(r_i, \mathbf{w}_i), i = 1, \dots, n : r_i > r_0\}$ , where  $r_0 > 0$  is a large threshold, comes approximately from a Poisson point process with intensity measure  $\psi$ . Let  $\mathcal{W}_{r_0} = \{(r, \mathbf{w}) : r > r_0\}$  be the set of points with a radial component larger than  $r_0$ , then the number of points falling in  $\mathcal{W}_{r_0}$  is given by  $N(\mathcal{W}_{r_0}) \sim \text{Pois}\{1/\psi(\mathcal{W}_{r_0})\}$ . Conditionally to  $N(\mathcal{W}_{r_0}) = m$ , the points  $\{(r_{(i)}, \mathbf{w}_{(i)}), i =$

$1, \dots, m\}$  are i.i.d. with common density  $d\psi(r, \mathbf{w})/\psi(\mathcal{W}_{r_0})$ . If we assume that  $H$  is known apart from a vector of unknown parameters  $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ , then the approximate likelihood of the excess is

$$\begin{aligned} L(\boldsymbol{\theta}; (r_{(i)}, \mathbf{w}_{(i)}), i = 1, \dots, m) &= \frac{e^{-\psi(\mathcal{W}_{r_0})} \psi(\mathcal{W}_{r_0})^m}{m!} \prod_{i=1}^m \frac{d\psi(r_{(i)}, \mathbf{w}_{(i)})}{\psi(\mathcal{W}_{r_0})} \\ &\propto \prod_{i=1}^m h(\mathbf{w}_{(i)}, \boldsymbol{\theta}), \end{aligned} \quad (21)$$

where  $h$  is a parametric angular density function (e.g. Engelke et al., 2015; Beirlant et al., 2006, pp. 170–171). In the next section the angular density models described in Section 3 are fitted to the data by the maximization of the likelihood (21). For brevity the asymmetric logistic model is not considered since it has too many parameters. The likelihood (21) is proportional to the product of angular densities, therefore the maximizer of (21) is obtained equivalently by maximizing the log-likelihood

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^m \log h(\mathbf{w}_{(i)}, \boldsymbol{\theta}). \quad (22)$$

Denote by  $\hat{\boldsymbol{\theta}}$  the maximizer of  $\ell$  and by  $\ell'(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta})$  the score function. Since (21) provides an approximation of the true likelihood, then from the theory on model misspecification (e.g. Davison, 2003, pp. 147–148) it follows that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}_p(\mathbf{0}, J(\boldsymbol{\theta})^{-1} K(\boldsymbol{\theta}) J(\boldsymbol{\theta})^{-1}), \quad n \rightarrow \infty,$$

where  $\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$  is the  $p$ -dimensional normal distribution with mean  $\boldsymbol{\mu}$  and covariance  $\Sigma$ ,  $\boldsymbol{\theta}$  is the true parameter and

$$J(\boldsymbol{\theta}) = -E\{\nabla_{\boldsymbol{\theta}} \ell'(\boldsymbol{\theta})\}, \quad K(\boldsymbol{\theta}) = \text{Var}_{\boldsymbol{\theta}}\{\ell'(\boldsymbol{\theta})\},$$

are the sensitive and variability matrices (Varin et al., 2011). In the case of misspecified models, model selection can be performed by computing the Takeuchi Information Criterion (TIC) (e.g. Sakamoto et al., 1986), that is

$$TIC = -2 \left[ \ell(\hat{\boldsymbol{\theta}}) - \text{tr}\{K(\hat{\boldsymbol{\theta}})J^{-1}(\hat{\boldsymbol{\theta}})\} \right],$$

where the log-likelihood, the variability and sensitive matrices are evaluated at  $\hat{\boldsymbol{\theta}}$ . The model with the smallest value of the  $TIC$  is preferred.

In order to derive an approximate posterior distribution for the parameters of an angular density,

the approximate likelihood (21) can be used within the Bayesian paradigm (see Sabourin et al., 2013). Briefly, let  $q(\boldsymbol{\theta})$  be a prior distribution on  $\boldsymbol{\theta}$ , then the posterior distribution of the angular density's parameters is

$$q(\boldsymbol{\theta}|\mathbf{w}) = \frac{\prod_{i=1}^m h(\mathbf{w}_{(i)}, \boldsymbol{\theta}) q(\boldsymbol{\theta})}{\int_{\Theta} \prod_{i=1}^m h(\mathbf{w}_{(i)}, \boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \quad (23)$$

With the angular density models in Section (3) the analytical expression of  $q(\boldsymbol{\theta}|\mathbf{w})$  can not be derived. Therefore, we use a Markov Chain Monte Carlo method for sampling from an approximation of  $q(\boldsymbol{\theta}|\mathbf{w})$ . Specifically, we use a Metropolis–Hastings simulating algorithm (e.g. Hastings, 1970). With the pairwise beta models we use the prior distributions described by Sabourin et al. (2013). With the titled Dirichlet and Hüsler-Reiss model we use independent zero-mean normal prior distributions with standard deviations equal to 3 for  $\log \alpha_j$  and  $\log \lambda_{i,j}$  with  $i, j \in I$ . For the extremal- $t$  model we use independent zero-mean normal prior distributions with standard deviations equal to 3 for  $\text{sign}(\rho_{ij})\text{logit}(\rho_{ij}^2)$  with  $i, j \in I$ , where  $\text{sign}(x)$  is the sign of  $x$  for  $x \in \mathbb{R}$  and  $\text{logit}(x) = \log(x/(1-x))$  for  $0 \leq x \leq 1$ , and a zero-mean normal prior distribution with standard deviations equal to 3 for  $\log \nu$ . Similar to Sabourin et al. (2013), for each models' parameter we select a sample of  $50 \times 10^3$  observations from the approximate posterior, after a burn-in period of length  $30 \times 10^3$ . These sizes have been determined using the Geweke convergence diagnostics (Geweke, 1992) and the Heidelberger and Welch test (Heidelberger and Welch, 1981) respectively.

Model selection is performed using the Bayesian Information Criterion (BIC) (e.g. Sakamoto et al., 1986), that is

$$BIC = -2\ell(\hat{\boldsymbol{\theta}}) + p\{\log m + \log(2\pi)\},$$

where  $p$  is the number of parameters and  $m$  is the sample size. The model with the smallest value of the BIC is preferred.

## 5 Real data analysis: Air quality data

We analyze the extremal dependence of the air quality data, recorded in the city centre of Leeds, UK. The aim is to estimate the probability that multiple pollutants will be simultaneously high in the near future. This dataset has been previously studied by Heffernan and Tawn (2004), Boldi and Davison (2007) and Cooley et al. (2010). The data are the daily maximum of five air pollutants: particulate matter (PM10), nitrogen oxide (NO), nitrogen dioxide (NO2), ozone (O3), and sulfur

dioxide (SO<sub>2</sub>). Levels of the gases are measured in parts per billion, and those of PM<sub>10</sub> in micrograms per cubic meter. We focus our analysis on the winter season (from November to February) from 1994 to 1998.

A preliminary analysis focuses on the data of triplets of variables. For brevity we only report the results of the most dependent triplets: PM<sub>10</sub>, NO, SO<sub>2</sub> (PNS), NO<sub>2</sub>, SO<sub>2</sub>, NO (NSN) and PM<sub>10</sub>, NO, NO<sub>2</sub> (PNN). For each variable, the empirical distribution function is estimated with the data below the 0.7 quantile and a GPD is fitted to the data above the quantile (Cooley et al., 2010). Then, each marginal distribution is transformed into a unit Fréchet. The coordinates of the data-points are transformed to radial distances and angular components. For each triplet, the 100 observations with the largest radial distances are retained. The angular density models in Section 3 are fitted to the data using the methods in Section 4.

The results are presented in Table 1. Maximum likelihood estimates are similar to the estimated posterior means and the estimated posterior standard deviations are typically larger than the standard errors. For PNS we obtain the same maximum likelihood estimates as Cooley et al. (2010) with the pairwise beta model, however we use (4) to compute the variances of the estimates and so we attain larger standard errors than they do. Both the TIC and BIC lead to the same model selection. The Hüsler-Reiss model provides the best fit for all the groups of pollutants.

From top to bottom, Figure 3 displays the angular densities, computed with the posterior means. From left to right the Hüsler-Reiss, the tilted Dirichlet and the pairwise beta densities are reported. With PNS, we see that there are many observations along the edge that link PM<sub>10</sub> and NO, revealing strong dependence between these two pollutants. There are also several observations on the SO<sub>2</sub> vertex, reflecting that this pollutant is mildly dependent with the other two. There are also some data in the middle of the simplex, indicating that there is mild dependence among the pollutants. Similarly, with NSN we see that there is strong dependence between NO and NO<sub>2</sub>, because there are many observations along the edge that link them. There is a mild dependence between SO<sub>2</sub> and the other pollutants, because there is a considerable amount of data on the O<sub>3</sub> vertex. Overall, there is mild dependence among the pollutants, because there is a small amount of data in the middle of the simplex. With PNN we see that most of the observations are placed on the middle of the simplex revealing an overall strong dependence among the pollutants. There is a small amount of data along the edge that link NO<sub>2</sub> and NO and on the PM<sub>10</sub> vertex. This reflects more dependence between NO<sub>2</sub> and NO than between NO<sub>2</sub> and PM<sub>10</sub> and PM<sub>10</sub> and NO. All these features are well captured by the angular densities estimated using the Hüsler-Reiss model.



Model Method Estimates					$\ell(\hat{\theta})$	TIC/BIC
TD		$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$		
PNS	L	1.20(0.24)	0.67(0.07)	0.41(0.08)	199.63	-399.21
	B	1.22(0.25)	0.68(0.11)	0.42(0.09)		-379.90
NSN	L	0.85(0.12)	0.39(0.08)	0.90(0.11)	200.84	-401.63
	B	0.86(0.15)	0.39(0.09)	0.81(0.15)		-382.32
PNN	L	1.43(0.28)	1.55(0.31)	1.28(0.20)	186.35	-372.64
	B	1.45(0.30)	1.57(0.28)	1.29(0.23)		-353.36
PB		$\hat{\beta}_{1,2}$	$\hat{\beta}_{1,3}$	$\hat{\beta}_{2,3}$	$\hat{\alpha}$	
PNS	L	3.21(0.70)	0.47(0.05)	0.45(0.04)	0.68(0.06)	95.95 -191.87
	B	3.31(1.13)	0.48(0.11)	0.46(0.10)	0.68(0.09)	-166.10
NSN	L	0.40(0.03)	3.74(1.77)	0.50(0.05)	0.64(0.05)	102.59 -205.13
	B	0.40(0.09)	4.00(1.72)	0.51(0.12)	0.64(0.08)	-179.36
PNN	L	3.75(1.38)	0.71(0.09)	3.18(1.21)	1.35(0.18)	84.31 -168.55
	B	3.83(1.75)	0.72(0.16)	3.70(1.80)	1.37(0.20)	-142.66
HR		$\hat{\lambda}_{1,2}$	$\hat{\lambda}_{1,3}$	$\hat{\lambda}_{2,3}$		
PNS	L	0.65(0.06)	0.90(0.04)	0.98(0.03)	234.51	-468.93
	B	0.65(0.04)	0.90(0.04)	0.98(0.04)		-449.67
NSN	L	1.00(0.04)	0.56(0.04)	0.96(0.04)	251.80	-503.54
	B	1.00(0.04)	0.57(0.03)	0.97(0.04)		-484.25
PNN	L	0.60(0.05)	0.70(0.04)	0.51(0.03)	198.23	-396.38
	B	0.60(0.03)	0.70(0.04)	0.51(0.03)		-377.11
ET		$\hat{\rho}_{1,2}$	$\hat{\rho}_{1,3}$	$\hat{\rho}_{2,3}$	$\hat{\nu}$	
PNS	L	0.87(0.02)	0.74(0.03)	0.66(0.03)	3.89(0.51)	152.13 -304.18
	B	0.87(0.02)	0.77(0.02)	0.72(0.01)	4.02(0.35)	-275.13
NSN	L	0.58(0.04)	0.87(0.02)	0.64(0.03)	3.50(0.01)	141.92 -283.80
	B	0.72(0.01)	0.89(0.02)	0.73(0.02)	4.00(0.33)	-242.50
PNN	L	0.88(0.02)	0.82(0.02)	0.89(0.01)	3.70(0.78)	180.74 -361.38
	B	0.86(0.02)	0.78(0.03)	0.87(0.02)	3.21(0.43)	-330.33

Table 1: Summary of the extremal dependence models fitted to the UK air pollution data. For each angular density model the estimation results of the triplets of pollutants are reported. L and B denote the approximate likelihood and Bayesian inferential method. Estimates are maximum likelihood (standard errors) and posterior means (standard deviations).

With this analysis we found that O3 is only weakly dependent with the other pollutants. This result was also found by Heffernan and Tawn (2004). Then, the second part of the analysis focuses only on PM10, NO, NO2 and SO2. Now, because a larger number of parameters needs to be estimated, then the 200 observations with the largest radial distances are selected (see Cooley et al., 2010). Table 2 presents the estimation results. For brevity we only report the maximum value of the log-likelihood, the TIC and the BIC. The Hüsler-Reiss model provides the smallest values of the TIC and BIC, revealing again that it better fits the pollution data. Accordingly hereafter calculations will be made using this model and the estimates obtained with the Bayesian

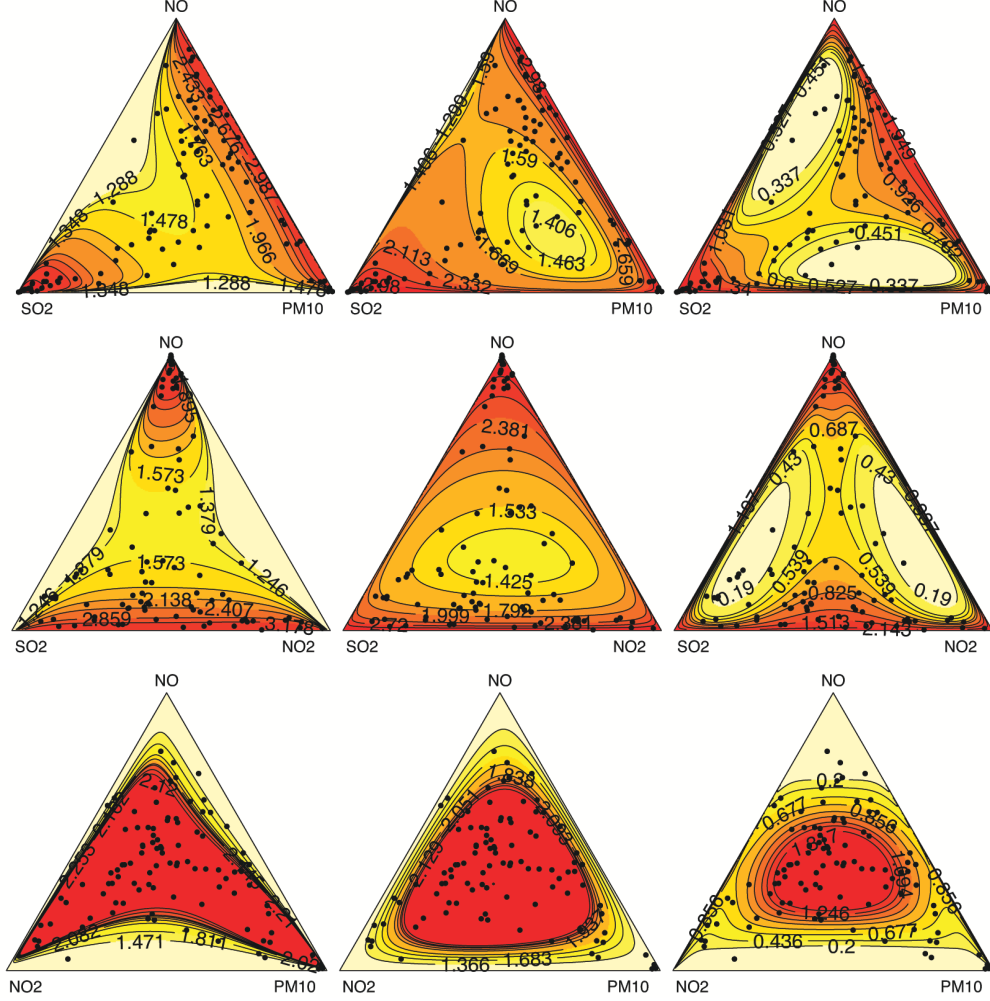


Figure 3: Estimated angular densities in logarithm scale. Dots represent the largest 100 observations.

	Tilted Dirichlet	Pairwise Beta	Hüsler-Reiss	extremal- $t$
$\ell(\hat{\theta})$	654.3	402.5	762.7	532.3
TIC	-1308.6	-805.0	-1525.3	-1064.5
BIC	-1280.0	-753.4	-1475.5	-974.7

Table 2: Summary of the extremal dependence models fitted to the UK air pollution data.

approach.

We summarize the extremal dependence of the four variables using the extremal coefficient (12) and the coefficient of tail dependence (15). Specifically,  $\hat{\vartheta} = 2.267$  with a 95% credible interval is equal to (1.942, 2.602) and  $\hat{\chi} = 0.242$  with a 95% credible interval is (0.150, 0.361). These results suggest a strong extremal dependence among the pollutants. The estimated extremal dependence can be used in turn to estimate the probability that multiple pollutants exceed a high threshold.

	Event 1	Event 2	Event 3
Excess / $n$	18/528	14/562	12/528
Emp. Est.	0.034 (0.019, 0.050)	0.025 (0.012, 0.038)	0.023 (0.010, 0.035)
Mod. Est.	0.038	0.030	0.030

Table 3: Probability estimates of excesses. The first row reports the number of excess and the sample size. The second row reports the empirical estimates and between brackets the 95% confidence intervals obtained with the normal approximation. The third row reports the model estimates.

Consider a value  $\mathbf{y}$  whose radial component is a high threshold  $r_0$ . Then, the probability of falling in the failure region (13) is approximately equal to the right hand side of (17). Because the exponent function is related to the tail function by the inclusion-exclusion principle, then using (19) we have

$$\text{pr}\{Y_1 > y_1, \dots, Y_d > y_d\} \approx \sum_{j=1}^d \frac{1}{y_j} \bar{\Phi}_{d-1} \left\{ \left( \lambda_{k,j} + \frac{\log y_k / y_j}{2\lambda_{k,j}} \right)_{k \in I_j}; \bar{\Lambda}_j \right\}, \quad (24)$$

where  $\bar{\Phi}_{d-1}$  is the survival function of the multivariate normal distribution (Nikoloulopoulos et al., 2009). Similar to Cooley et al. (2010) we define three extreme events:  $\{\text{PM10} > 95, \text{NO} > 270, \text{SO2} > 95\}$ ,  $\{\text{NO2} > 110, \text{SO2} > 95, \text{NO} > 270\}$  and  $\{\text{PM10} > 95, \text{NO} > 270, \text{NO2} > 110, \text{SO2} > 95\}$ . Then, we compute probability (24) using in place of the parameters their estimates. Table 3 reports the results. For the three events the estimates fall inside the 95% confidence intervals highlighting the ability of the model to estimate such extreme events.

The right-hand side of (24) can also be used for estimating joint return levels. In the univariate case see Coles (2001, pp.49–50). In the multivariate case different definitions of return levels may be available (Johansen, 2004). Let  $J \subset I$ ,  $\{x_i, i \in I \setminus J\}$  be a sequence of fixed high thresholds and  $p \in (0, 1)$  be a fixed probability. Given a return period  $1/p$ , we define *joint return levels* the quantiles  $\{y_{j;p}, j \in J\}$  that satisfy the equation

$$p = \text{pr}(Y_j > y_{j;p}, Y_i > x_i, j \in J, i \in I \setminus J).$$

Figure 4 displays univariate and bivariate *joint return level plots*. When  $J = \{j\}$ , with  $j \in I$ , then the joint return level plot displays  $y_{j;p}$  against  $1/p$  for different values of  $p$ . When  $J = \{i, j\}$ , with  $i, j \in I$ , then for different values of  $1/p$  the contour levels of  $(y_{i;p}, y_{j;p})$  are displayed. With solid lines, the top-left and right panels of Figure 4 report the estimated return levels of NO2 and PM10 jointly to the extreme events  $\{\text{SO2} > 95, \text{NO} > 270\}$  and  $\{\text{NO} > 270, \text{NO2} > 110, \text{SO2} > 95\}$  respectively. The dots are the empirical estimates and the red solid lines are the pointwise 95%

confidence intervals. These are computed using the normal approximation when  $p > 0.02$  and using exact binomial confidence intervals when  $p < 0.02$ . The bottom-left and right panels report the contour levels of the return levels for (NO<sub>2</sub>, SO<sub>2</sub>) and (PM<sub>10</sub>, NO) jointly to events  $\{\text{NO} > 270\}$  and  $\{\text{NO}_2 > 110, \text{SO}_2 > 95\}$  respectively.

The joint return level can be interpreted as follows. For example, from the top-right panel we have that the 50 years joint return level of PM<sub>10</sub> is 166. Concluding, we expect that PM<sub>10</sub> will exceed the level 166 together with the event that NO, SO<sub>2</sub> and NO<sub>2</sub> simultaneously exceed the levels 270, 95 and 110 respectively, on average every 50 years.

## 6 Computational details

The figures and the estimation results have been obtained using the free software R (Team, 2013) and in particular the package **ExtremalDep**, available at <https://r-forge.r-project.org/projects/extremaldep/>. Bayesian estimation is obtained using and extending some routines of the package **BMamev**. The left and middle panels of Figure 1 were obtain using the routines `scatter3d` and `polygon3d` of the

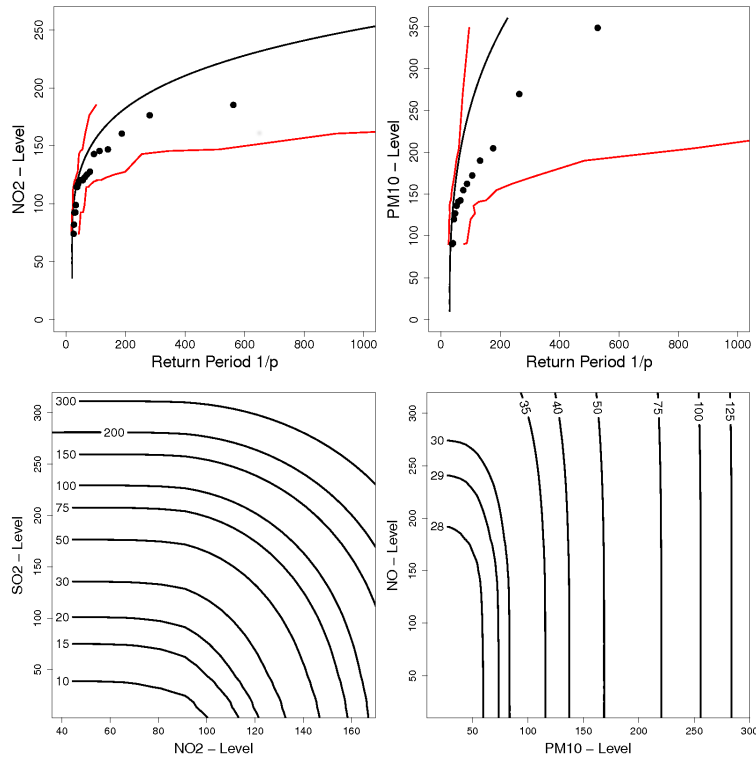


Figure 4: Joint return level plots of single components NO<sub>2</sub> and PM<sub>10</sub> and of the two components (SO<sub>2</sub>, NO<sub>2</sub>) and (NO, PM<sub>10</sub>)

package `plot3D`.

## References

- Apputhurai, P. and Stephenson, A. (2011), “Accounting for uncertainty in extremal dependence modeling using bayesian model averaging techniques,” *Journal of Statistical Planning and Inference*, 141, 1800–1807.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2006), *Statistics of extremes: theory and applications*, John Wiley & Sons.
- Boldi, M.-O. and Davison, A. C. (2007), “A mixture model for multivariate extremes,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 69, 217–229.
- Coles, S. (2001), *An introduction to statistical modeling of extreme values*, Springer Series in Statistics, Springer-Verlag London, Ltd., London.
- Coles, S. G. and Tawn, J. A. (1991), “Modelling Extreme Multivariate Events,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 53, pp. 377–392.
- (1994), “Statistical Methods for Multivariate Extremes: An Application to Structural Design,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43, pp. 1–48.
- Cooley, D., Davis, R. A., and Naveau, P. (2010), “The pairwise beta distribution: a flexible parametric multivariate model for extremes,” *J. Multivariate Anal.*, 101, 2103–2117.
- Davison, A. C. (2003), *Statistical Models*, Cambridge.
- Davison, A. C. and Gholamrezaee, M. M. (2012), “Geostatistics of extremes,” *Proceedings of the Royal Society of London Series A: Mathematical and Physical Sciences*, 468, 581–608.
- Davison, A. C., Padoan, S. A., and Ribatet, M. (2012), “Statistical Modeling of Spatial Extremes,” *Statistical Science*, 27, 161–186.
- de Haan, L. and Ferreira, A. (2006), *Extreme value theory*, Springer Series in Operations Research and Financial Engineering, Springer, New York, an introduction.
- de Haan, L. and Resnick, S. I. (1977), “Limit theory for multivariate sample extremes,” *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40, 317–337.

- De Haan, L. and Zhou, C. (2011), “Extreme residual dependence for random vectors and processes,” *Advances in Applied Probability*, 43, 217–242.
- Demarta, S. and McNeil, A. J. (2005), “The t Copula and Related Copulas,” *International Statistical Review*, 73, 111–129.
- Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015), “Estimation of Huesler-Reiss distributions and Brown-Resnick processes,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77, 239–265.
- Falk, M., Hüsler, J., and Reiss, R.-D. (2011), *Laws of small numbers: extremes and rare events*, Birkhäuser/Springer Basel AG, Basel, extended ed.
- Geweke, J. (1992), *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, vol. Bayesian Statistics, University Press.
- Gudendorf, G. and Segers, J. (2011), “Nonparametric estimation of an extreme-value copula in arbitrary dimensions,” *Journal of Multivariate Analysis*, 102.
- (2012), “Nonparametric estimation of multivariate extreme-value copula,” *Journal of Statistical Planning and Inference*, 142.
- Gumbel, E. J. (1960), “Distributions des valeurs extremes en plusieurs dimensions,” *Publ. Inst. Statist. Univ. Paris*, 9.
- Hastings, W. K. (1970), “Monte Carlo Sampling Methods Using Markov Chains and Their Applications,” *Biometrika*, 57, pp. 97–109.
- Heffernan, J. E. and Tawn, J. A. (2004), “A conditional approach for multivariate extreme values,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66, 497–546, with discussions and reply by the authors.
- Heidelberger, P. and Welch, P. D. (1981), “A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations,” *Commun. ACM*, 24, 233–245.
- Hüsler, J. and Reiss, R.-D. (1989), “Maxima of normal random vectors: between independence and complete dependence,” *Statist. Probab. Lett.*, 7, 283–286.
- Joe, H. (1997), *Multivariate models and dependence concepts*, vol. 73 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London.

- Johansen, S. S. (2004), “Bivariate frequency analysis of flood characteristics in Glomma and Gudbrandsdalslagen,” Ph.D. thesis, University of Oslo Department of Geosciences Section of Geohazards and Hydrology.
- Kotz, S. and Nadarajah, S. (2000), *Extreme value distributions*, London: Imperial College Press, theory and applications.
- Ledford, A. W. and Tawn, J. A. (1996), “Statistics for near independence in multivariate extreme values,” *Biometrika*, 83, 169–187.
- (1997), “Modelling dependence within joint tail regions,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59, 475–499.
- Li, H. (2009), “Orthant tail dependence of multivariate extreme value distributions,” *Journal of Multivariate Analysis*, 100, 243–256.
- Marcon, G., Padoan, S. A., Naveau, P., and Muliere, P. (2014), “Multivariate Nonparametric Estimation of the Pickands Dependence Function using Bernstein Polynomials,” *Under review*.
- Nikoloulopoulos, A. K., Joe, H., and Li, H. (2009), “Extreme value properties of multivariate  $t$  copulas,” *Extremes*, 12, 129–148.
- Nolan, J. (2003), *Stable distributions: models for heavy-tailed data*, Birkhauser.
- Padoan, S. A. (2013a), *Encyclopedia of Environmetrics*, John Wiley & Sons, Ltd., Chichester, chap. Max-Stable Processes, no. 4.
- (2013b), *Encyclopedia of Environmetrics*, John Wiley & Sons, Ltd., Chichester, chap. Extreme Value Analysis, no. 2.
- (2013c), “Extreme Dependence Models Based on Event Magnitude,” *Journal of Multivariate Analysis*, 122, 1–19.
- Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010), “Likelihood-Based Inference for Max-Stable Processes,” *Journal of the American Statistical Association*, 105, 263–277.
- Pickands, III, J. (1981), “Multivariate extreme value distributions,” in *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)*, vol. 49, pp. 859–878, 894–902, with a discussion.

- Resnick, S. I. (2007), *Extreme values, regular variation, and point processes*, Springer.
- Ribatet, M. (2013), “Spatial extremes: Max-stable processes at work,” *Journal de la Société Française de Statistique*, 154, 156–177.
- Sabourin, A. and Naveau, P. (2014), “Bayesian Dirichlet mixture model for multivariate extremes: A re-parametrization,” *Computational Statistics and Data Analysis*, 71, 542–567.
- Sabourin, A., Naveau, P., and Fougères, A.-L. (2013), “Bayesian model averaging for multivariate extremes,” *Extremes*, 16, 325–350.
- Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986), *Akaike Information Criterion Statistics*, D. Reidel Publishing Company.
- Smith, R. L. (1990), “Max-stable processes and spatial extremes,” Tech. rep., University of Surrey.
- Tawn, J. A. (1990), “Modelling Multivariate Extreme Value Distributions,” *Biometrika*, 77, pp. 245–253.
- Team, R. D. C. (2013), *R: a Language and Environment for Statistical Computing*.
- Varin, C., Reid, N., and Firth, D. (2011), “An overview of composite likelihood methods,” *Statist. Sinica*, 21, 5–42.
- Wadsworth, J. L. and Tawn, J. A. (2012), “Dependence modelling for spatial extremes,” *Biometrika*, 1, 1–20.